

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

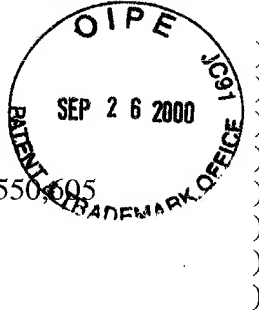
In re Application of:

Leif ANDERSSON et al.

Application Number: 09/550,695

Filed: April 17, 2000

For: METHODS FOR DETERMINING COAT COLOUR GENOTYPES IN PIGS



Group Art Unit: 1633

Examiner: To be Assigned

CLAIM FOR PRIORITY UNDER 35 U.S.C. § 119

BOX MISSING PARTS
Commissioner for Patents
Washington, D.C. 20231

Sir:

Applicants are enclosing a certified copy of Great Britain Patent Application No. 9722027.1, filed in Great Britain on October 17, 1997. This document provides a basis for Applicants' claim for priority.

No fee is believed due as a result of this submission. However, if a fee is due upon the filing of this priority document, please charge the undersigned's Deposit Account No. 02-0375.

Respectfully submitted,

BAKER BOTTS L.L.P.

By:

James B. Arpin
Registration No. 33,470

Dated: September 26, 2000
Baker Botts L.L.P.
The Warner; Suite 1300
1299 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2400
(202) 639-7700 (telephone)
(202) 639-7890 (facsimile)

JBA:nej



The
Patent
Office



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

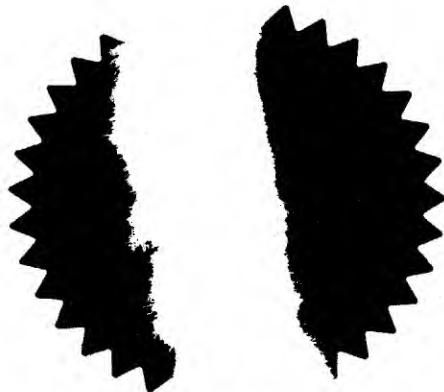


I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated

20 APR 2000

Patents Form 1/77

Patents Act 1977
(Rule 16)

The
Patent
Office



200CT97 E310978-13 D00056
P01/7700 25.00 - 9722027.1

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

PWC/P20434GB

2. Patent application number

(The Patent Office will fill in this part)

17 OCT 1997

9722027.1

3. Full name, address and postcode of the or of each applicant (underline all surnames)

MELICA HB
BERGAGATAN 30
S-752 39 UPPSALA
SWEDEN

✓
Patents ADP number (if you know it)

7221575001

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention

METHODS

5. Name of your agent (if you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

KILBURN & STRODE
30 JOHN STREET
LONDON
WC1N 2DD

Patents ADP number (if you know it)

125001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an applicant or

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-
Description	23
Claim(s)	3
Abstract	-
Drawing(s)	7

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

☒ Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature

Date

17th October 1997

12. Name and daytime telephone number of person to contact in the United Kingdom

Mr. Paul Chapman
Tel: 0171 242 8291

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.

METHODS

The present invention relates to methods for determining coat colour genotype in pigs. In particular, it relates to methods of distinguishing between the alleles I , I^P , I^* and i of the *KIT* gene.

5

Coat colour is important to the pig breeding industry for a number of reasons. Firstly in a number of markets there is a preference for white skinned meat. This is due to the fact that pork is often marketed with the skin still attached, and skins from coloured pigs, even if dehaired, can still exhibit coloured hair roots, which can lead to negative perception by the consumer, since the surface of the meat may appear to be spotted by mould. It is therefore necessary in these markets to remove the skin from such carcasses, entailing additional cost.

10

For example, in the US, coloured carcasses are associated with approximately 1% of skin defects requiring dehairing and skinning to remove pigment. As a result of this, coloured pig carcasses are generally discounted. Secondly gross variation in the appearance of pigs claimed to be genetically consistent for other traits can lead to questions about the consistency and quality of the animals in the mind of pig-producing customers. Breeders would also like to be able to ensure consistency in breeding populations. Thus breeders may wish to ensure that progeny produced by breeding crosses were always white. Alternatively a breeder producing a coloured breed may wish to ensure that the correct coat colour characteristics were maintained even during the introgression of genes from white lines.

15

20

White coat colour in pigs is controlled by the dominant white locus designated I (for inhibition of coat colour). Structural alterations in the porcine *KIT* gene have recently been correlated with various alleles of I and are probably responsible for the differences in coat colour pattern found or are closely linked to the mutations that are

25

found (Johansson Moller *et al.* 1996 Mammalian Genome 7, 822-830). Four structural versions of the porcine *KIT* gene have been identified to date and designated *I*, *I^P*, *I^{*}* and *i* (see figure 1 and table 1). The version found in fully coloured animals including wild boar and which is therefore generally accepted as the wild type allele is *i*. The other versions of the gene known all involve duplication of at least part of the porcine *KIT* gene. *I^P* is a partially dominant allele and causes in the heterozygous state (*I^P/i*) a phenotype "patch" characterised by patches of white and coloured coat. *I* and *I^{*}* are both fully dominant and cause a white phenotype both in the heterozygous and homozygous state. The only difference between *I* and *I^{*}* is that there is a 4bp deletion in intron 18 in one of the two *KIT* gene copies associated with the *I* allele. This sequence polymorphism is not expected to have any functional effect.

The phenotypes of a number of gene combinations are listed below with the *I* allele being dominant over *I^P* and *i*, and *I^P* being dominant over *i* (Johansson *et al.* 1992 Genomics 14, 965-969)

<u>Genotype</u>	<u>Colour</u>
<i>I/I</i>	White
<i>I/I^P</i>	White
<i>I/i</i>	White
<i>I^P/i</i>	Patched
<i>i/i</i>	Coloured
<i>I[*]/i</i>	White

Previously Moller Johansson *et al.* (1996 Mammalian Genome 7, 822-830) revealed that *KIT* occurs as a single copy gene in coloured (*i/i*) animals but is duplicated in the *I*, *I^{*}* and *I^P* allele. This duplication allowed the differentiation of *I*, *I^{*}* and *I^P* from *i*

through examination of the gene copy number of KIT and to use linked polymorphisms in or in the near vicinity of the KIT gene to distinguish different alleles at the *I* locus. This approach is the subject of International patent application No. PCT/GB96/01794.

5

A problem remaining from this method however is that it does not distinguish between *I*^{*} and *I*^P. The result of this is that if one wishes to use the screen to remove heterozygous carrier animals of genotype *I/I*^P from a white pig line one also excludes *I/I*^{*} animals unnecessarily. The removal of animals, potentially very valuable if at the top of a breeding pyramid, unnecessarily can lead to a reduction in the rate of improvement of other traits within that particular group of animals. Other consequences include a general loss of genetic diversity and a loss of alleles at the locus in question that may have as yet undetected value. It is essential that one only excludes alleles where absolutely necessary.

10

15

Further sequence analysis of genomic DNA revealed a mutation at the first nucleotide in intron 17 of KIT2 leading to a defect in the splicing of RNA transcribed from this particular copy of the KIT gene. In experiments involving several pig breeds, this mutation showed a complete concordance with the presence of the *I* and *I*^{*} alleles but was not found in the *i* and *I*^P alleles

20

The majority of eukaryotic genes are composed of both coding (exon) and intervening noncoding (intron) regions. The latter of these sequences are removed from large pre-mRNA by a highly accurate cleavage and ligation reaction known as splicing. The result is a mature mRNA transcript, devoid of intron sequences, which is transported to the cytoplasm for translation. The splicing of eukaryotic genes is most likely a two step procedure. First, the pre-mRNA is cleaved at the 5' (donor) splice site followed by cleavage at the 3' (acceptor) splice site. The second step involves rejoining of the spliced exons to result in exclusion of the intervening intron. Splicing is therefore

25

critically dependant on the accuracy of the cleavage and ligation reactions. This accuracy appears to be dependant on the almost completely invariant GT and AG dinucleotides present at the 5' and 3' exon/intron boundaries respectively. These dinucleotides and the often highly conserved surrounding sequences are known as splice sites and serve to bind the protein factors required to perform the cleavage and ligation reactions.

The consequences for mutation occurring within a splice site may be a reduction in the amount of mature mRNA produced and/or utilization of alternative but incorrect splice sites in the vicinity. The result is production of mRNA which either contains additional intron sequence or which may lack a portion of coding sequence. Where mutation occurs within the 5' (donor) splice site and prevents binding of protein factors, the exon is no longer recognised as such and is excised along with its neighbouring introns. This is referred to as 'exon skipping' (as reviewed by Cooper and Krawczak, 1994 Human Gene Mutation. Bios Scientific Publishers, Oxford, UK, 1994).

In the mutation identified here the G of the conserved GT pair at the exon17/intron17 boundary region is altered to an A. That there is an alteration in the messenger RNA that is translated into protein and that the presence of the change correlates with the white/patched phenotypes suggests that this is the functional mutation rather than merely a linked marker. The splice variant is expected to give rise to a defective protein as 41 amino acids in the mature protein are missing. We therefore assume that this splice mutation is the causal mutation for the difference between *I* and *I^P*. Our current knowledge as regards molecular differences between alleles at the *I* locus is summarised in Table 1. In conclusion, we have identified two functionally important mutations. One is the gene duplication present in *I^P*, *I^{*}* and *I* which by itself appears to cause the patch phenotype. The exact reason for this phenotypic effect is not known but one can speculate on the basis of comparative data from the mouse that the

duplicated copy of the KIT gene may lead to a defect in gene expression which in turn affects melanocyte migration. This is especially valuable in that there can be no breakdown of the linkage between the DNA polymorphism and the trait itself. This has allowed us to develop a range of assays for the determination of the presence of the DNA polymorphism and the genotype of the animal with regard to coat colour determination to a significantly greater extent than has been possible before. The second mutation is the splice mutation that occurs in one of the KIT copies associated with the dominant white (*I* and *I**) alleles. The expression of a truncated form of the KIT protein is expected to cause a more severe defect in KIT function and a more severe effect on coat colour.

Table 1.

Allele	KIT gene	Intron 17	Associated Phenotype
<i>i</i>	KIT1	Normal	Coloured
<i>I^p</i>	KIT1 KIT2	Normal Normal	Patch
<i>I</i>	KIT1 KIT2	Normal Mutated	Dominant White
<i>I*</i>	KIT1 KIT2	Normal Mutated	Dominant White

In addition to these two functionally important mutations, a mutation in the KIT gene with no known phenotypic effect, the 4 bp deletion in intron 18 has been documented at the *I* locus as described in International patent application No. PCT/GB96/01794.

Thus, in a first aspect the present invention provides a method for determining coat colour genotype in a pig which comprises:

- (a) obtaining a sample of pig nucleic acid; and

(b) analysing the nucleic acid obtained in (a) to determine whether a mutation is present at one or more exon/intron splice sites of the *KIT* gene.

In particular, the method determines whether a mutation is present at the exon
5 17/intron 17 boundary, eg the substitution of the G of the conserved GT pair for A.

Reverse Transcriptase based Polymerase Chain analysis (RT-PCR) of the exon 16-19
region of *KIT* mRNA in animals of the Large White breed (*III*) and comparison to that
transcribed in Hampshire animals (*i/i*) revealed an extra species of molecule in the
10 former animals. RT-PCR analysis of both breeds yielded a product fragment with a
length of 424bp indicating that both types of animal contained a *KIT* mRNA transcript
containing a region corresponding to this region of the gene. However in addition the
III animals yielded a RT-PCR product of 301bp indicating the presence of an mRNA
species which did not contain the full transcription of the exon 16-19 DNA sequence
15 (see example 1). These two transcripts have been shown to be derived from the
separate duplicate copies of the *KIT* gene associated with the *I* allele. Sequencing over
the *KIT* exon 17/intron 17 boundary revealed a difference in the intron boundary
sequences present in the two duplicate copies of the *KIT* gene associated with the *I*
allele (see example 2). The sequences are as shown below:

20	Allele	Kit Gene	Exon17 Intron 17
	<i>i</i>	KIT1AAC GTG....
	<i>I^P</i>	KIT1AAC GTG.....
	<i>I^P</i>	KIT2AAC GTG.....
	<i>I</i>	KIT1AAC GTG.....
25	<i>I</i>	KIT2AAC ATG.....

This alteration of the 5' intron splice site from a GT pair to an AT pair affects the splicing of the pre mRNA and results in the loss of the whole of exon 17 from the mRNA transcribed from the *I-KIT2* sequence. This will result in a modified KIT protein with the associated alterations in function and phenotype.

5

Based upon the sequence polymorphism rapid tests can be developed to determine the alleles carried by a specific animal at the dominant white locus.

10

In one form such a test would comprise amplification of the region through the polymerase chain reaction (PCR) utilising genomic DNA from the animal in question as a template. The nucleotide sequence CATG comprises a recognition sequence for the restriction enzyme *NlaIII*. This sequence is only present in the *I-KIT 2* sequence at the junction position and thus one can differentiate the DNA molecules amplified from the two alleles by digestion of the amplification products with this enzyme or any other restriction enzyme with a suitable recognition site.

15

Genomic DNA for use in such a test can be prepared by a wide range of available methods, from any tissues or products derived from the animal in question. A 175bp fragment of the *KIT* gene containing the exon 17/intron 17 boundary region can be amplified from porcine genomic DNA using a pair of primers such as:

20

KIT21 (5'-GTA TTC ACA GAG ACT TGG CGG C-3'); and

KIT35 (5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3').

25

The use of such primers in a PCR-RFLP test yields a fragment of 175bp before digestion with *NlaIII*. In alleles with the G present at nucleotide position 1 of intron 17

(*I-KIT1* type sequence) there is only one *Nla*III cleavage site present 41bp from the end of the fragment. This site is present in all versions of the *KIT* sequence. Thus digestion of the 175bp fragments obtained from *i* and *P* alleles yields two fragments of 134bp and 41bp. Where the G is mutated to A as in the *I-KIT2* sequence present in *I* and *I** a further *Nla*III cleavage site is created and thus digestion yields products of 80bp and 54bp and 41bp (see figure 3, example 2). A number of other oligonucleotides suitable as PCR primers could easily be derived from the sequence of this region of the porcine genome. An example of such a PCR-RFLP test is given in example 3. The results that would be obtained from such a PCR-RFLP test described above are as shown below:

	Genotype	Fragment sizes	Fragment sizes
		134 + 41 bp	80 + 54 + 41 bp
	<i>III</i>	Yes	Yes
	<i>I/P</i>	Yes	Yes
15	<i>I/i</i>	Yes	Yes
	<i>P/P</i>	Yes	No
	<i>P/i</i>	Yes	No
	<i>i/i</i>	Yes	No
	<i>I*/I</i>	Yes	Yes
20	<i>I*/I*</i>	Yes	Yes
	<i>i/I*</i>	Yes	Yes
	<i>I*/P</i>	Yes	Yes

Thus, simply by analysing restriction products certain genotypes can be distinguished.

However, there are others which cannot be distinguished in this first configuration of the test. In a further refinement the test can be carried out in such a way that the amount of each fragment can be calculated. By carrying out electrophoresis on an apparatus that allows the quantification of each of the bands one can determine the ratio of the two forms of template in the genomic DNA sample used. Examples of

such an apparatus include the Perkin Elmer Applied Biosystems 373 and 377 DNA sequencing systems. The application of this type of equipment is illustrated in example 4. Any equipment capable of determining the relative amounts of the products from the two different sequences is equally applicable to such tests. The expected results from such a test are shown below.

Genotype	Normal KIT sequence Copies	Splice mutant KIT Copies	Ratio <u>Normal</u> Splice mutant
<i>III</i>	2	2	1
<i>I/P</i>	3	1	3
<i>II/I*</i>	2	2	1
<i>II/i</i>	2	1	2
<i>P/P</i>	4	0	0
<i>P/I*</i>	3	1	3
<i>P/i</i>	3	0	0
<i>I*/i</i>	2	1	2
<i>I*/I*</i>	2	2	1
<i>i/i</i>	2	0	0

Thus, using such a test one could identify all animals carrying alleles of dominant white that might dispose themselves or their offspring to exhibiting non-white coat colour (*i* or *P*) as those giving a ratio other than one. Depending on the requirement and the derivation of the lines under selection one could take an appropriate subset of animals. For example in a cross derived animals carrying only alleles *I* and *i* one could identify any white individuals (*III* or *II/i*) carrying *i* as they would have a ratio in the test of 2 as opposed to 1 for the *III* animals. The distinction of alleles *I* and *I** can be carried out on the basis of the 4bp deletion as described in the previously filed patent PCT/GB96/01794.

There are a range of techniques by which differentiation of alleles containing the splice mutation and those containing the normal sequence could be differentiated by a person expert in the field.

- 5 Analysis of the genetic composition of an animal could be based upon a number of different source materials. These include genomic DNA, RNA and the KIT protein itself. There may also be effects on the levels and nature of other proteins, metabolites and RNA species which could be measured to create a more indirect assay.
- 10 DNA could be used as the basis for a number of approaches to testing. One approach is through the amplification of the region of DNA containing the polymorphism using the polymerase chain reaction. This could then be linked to a number of forms of analysis of the product. Examples of electrophoresis based technologies include Single Strand Conformation Polymorphism (SSCP), Restriction Fragment Length
- 15 Polymorphism (RFLP) and DNA sequencing of PCR products or direct genome sequencing. Other PCR based techniques that might alternatively be used include the Perkin Elmer TaqMan systems, Single Nucleotide Polymorphic Extension (SNuPE) and Minisequencing. PCR products from the region might also have application in hybridization based approaches to the differentiation of alleles at this locus.
- 20 Hybridization methods might include the probing of Southern transfers of genomic DNA with allele specific oligonucleotide, RNA, DNA fragment or Protein Nucleic Acid (PNA) probes. Other strategies of application here include hybridization of genomic DNA or PCR products (specific or whole genome) to oligonucleotide arrays possibly in the form of 'DNA chips'. Such arrays could consist of any reagent capable
- 25 of binding DNA or RNA derived from the locus in question in an allele specific manner. Further useful methods of analysis also include oligonucleotide ligation assay and the ligase chain reaction. For a review on methods for detecting point mutations see Landegren, 1996, Laboratory Protocols for Mutation detection, Oxford University press, Oxford.

A number of effects on the RNA produced from the gene in question have already and may in the future be observed. All such differences between the mutated and normal forms of the gene are useful targets for the determination of genotype and a large range of methods are available to the person skilled in the art. The changes that are or might be observed and methods of analysis are as follows. Alteration of the size, rate of processing, stability and quantity of RNA transcripts could be measured through widely used techniques such as northern blotting and RT-PCR as well as a number of the techniques described above for DNA analysis such as hybridization to oligonucleotide or DNA fragment arrays.

Another approach which can be used is to use a linked genetic polymorphism which is closely associated with the presence or absence of the alteration at the exon/intron boundary. Such a polymorphism may occur in the *KIT* gene itself or in a chromosomal region linked to *KIT*. By using a single linked marker in complete association with the presence/absence of the duplication or a combination of markers showing a partial association a highly informative test can be developed. For instance, the SSCP (Single Strand Conformation Polymorphism) method may be used to develop such polymorphism. The principle of the method is that double-stranded DNA, produced by PCR, is denatured into single-stranded DNA which is then separated by non-denaturing gel electrophoresis. Under non-denaturing conditions the single-stranded DNA forms a secondary structure due to intra-strand interaction but a proportion of the single-stranded DNA will renature and form double-stranded DNA. Two types of polymorphism may be revealed by this method. Firstly, a difference in nucleotide sequence between two alleles may influence the secondary structure of single-stranded DNA which is revealed as a difference in the mobility rate during electrophoresis. Secondly, a difference in nucleotide sequence often influences the mobility of the heteroduplex DNA (A heteroduplex is a double-stranded DNA molecule formed by two single-stranded molecules representing different alleles).

Association between genetic markers and genes responsible for a particular trait can be disrupted by genetic recombination. Thus, the closer the physical distance between the marker and the gene in question, the less likely it is that recombination will separate them.

5

It is also possible to establish linkage between specific alleles of alternative DNA markers and alleles of DNA markers known to be associated with a particular gene (e.g. the *KIT* gene discussed herein), which have previously been shown to be associated with a particular trait. Thus, in the present situation, taking the *KIT* gene, it would be possible, at least in the short term, to select for pigs with a particular coat colour, indirectly, by selecting for certain alleles of a *KIT* gene associated marker through the selection of specific alleles of alternative chromosome 8 markers. Examples of such markers known to be linked to the *KIT* gene on porcine chromosome 8 include genetic polymorphism in the *KIT* gene itself or in the closely linked genes for the α -subunit of platelet derived growth factor (PDGFRA) and albumin.

15

Particular genetic markers associated with the *KIT* gene are microsatellites. These are simple sequence repeats of 4, 3 or, more usually, 2 nucleotides, which occur essentially at random around the genome at approximately every 50,000 bases (about 60,000 microsatellites per haploid genome). Stuttering of DNA polymerase during replication and unequal crossing-over during recombination are thought to result in the loss or gain of repeat units. This means that microsatellites are usually polymorphic and can have several repeat length alleles.

20

Examples of linked microsatellite sequences include S0086 (Ellegren *et al*, *Genomics*, 16:431-439 (1993)) , S0017 (Coppieters *et al*, *Animal Genetics* 24: 163-170 (1993)), Sw527, Swr750 and SW916 (Rhorer *et al*, *Genetics*, 136:231-245 (1994)) . It would be possible to select indirectly for alleles of the *KIT* gene linked to coat colour using any of the above markers, or indeed any other linked markers on porcine chromosome 8.

25

Alterations in the level of the KIT protein could be measured using either specific antibodies for example in an ELISA system or on western blots or through the use of a range of biochemical techniques to measure the activity of the protein. Such tests could also be applied to other proteins or metabolites, the level or nature of which is altered by the presence of specific alleles at the locus. Different protein structures due to the presence of specific alleles could be identified through the use of structure specific antibodies. As with the DNA and RNA based methods all these protein methods could be applied in a quantitative fashion thus achieving the full discriminatory capabilities possible.

Kits could be produced for the specific analysis of the polymorphism described here alone and also with reagents allowing the combined analysis of the other polymorphisms previously reported and the subject of patent application PCT/GB96/01794.

The invention will now be described with reference to the following examples, which should not be construed as in any way limiting the invention. The examples refer to the figures in which:

Figure 1: is a schematic representation of the structure of the known porcine *KIT* alleles, where 4bp del'n refers to the 4bp deletion in intron 18 of one copy of the duplicated *KIT* gene DNA as reported by Moller *et al.* 1996 and Exon 17-A refers to the change of nucleotide 1 of intron 17 from a G as in the wild type allele to an A as reported in this patent;

Figure 2: shows an electropherogram (4% agarose Nusieve/Seakem 3:1; 100V for 80 min) showing RT-PCR products of *KIT* exon 16-19 with the primers KIT1F and KIT7R. The samples 1-3 and 4-6 are Swedish Large White and Hampshire pigs, respectively. The size difference between the 424 and 301 bp

fragments is due to lack of exon 17 in the latter fraction. The two upper bands of the Yorkshire pigs were interpreted as heteroduplexes (HD);

Figure 3: shows a 48 bp sequence comprising 21 bp of KIT exon 17 and 27 bp of KIT intron 17 where the position of the intron/exon border is marked with a vertical line, the splice site mutation (nt1^{G→A}) indicated with a vertical arrow and identical bases in alleles *I^P* and *i* are marked with a dot;

Figure 4: shows the results of *Nla*III PCR RFLP test used to detect the presence of a splice site mutation in intron 17 of the KIT gene. Figure 4A shows the position of two *Nla*III recognition sites within the PCR product amplified using primer pair KIT21 and KIT35. All distances are given in base pairs. Figure 4B shows the size of fragments which result following *Nla*III digestion of either normal *KIT* or splice mutant *KIT*. Figure 4C illustrates use of the PCR RFLP test. Lane 1 shows the KIT21/KIT35 amplified fragment undigested. Digestion was performed on PCR products amplified from, in Lane 2: a clone which contains the splice site mutation; Lane 3: a clone which contains the normal splice site sequence; Lane 4: genomic DNA from a coloured pig; Lane 5: genomic DNA from a white pig. Fragment sizes are given in base pairs;

Figure 5: shows a comparison of the ratio of normal to splice mutant *KIT* in animals of genotypes *I/I*, *I/i* and *I/I^P*; and

Figure 6: shows the ratio values for 56 Landrace and 33 Large White animals. A clearly bimodal distribution is observed with 7 Landrace and 3 Large White individuals having a ratio value of approximately 3 or above, suggesting them to be heterozygous carriers for the *I^P* allele (genotype *I/I^P*). This means *I^P* has gene frequency estimates of 6.25% (7/112 chromosomes tested) and 4.5%

(3/66 chromosomes tested) within the Landrace and Large White breeds respectively.

5 **Example 1**

RT-PCR of porcine *KIT* exon 16-19

i. mRNA purification from blood samples

10 Fresh blood samples were collected in citrate tubes from coloured Hampshire pigs and
 ✓ Large White pigs. Leukocytes were isolated from 5 ml blood using Ficoll 100
 (Pharmacia Biotech). Isolation of mRNA from leukocytes was then carried out using
 the Quickprep *Micro* mRNA purification kit (Pharmacia Biotech). The mRNA was
 stored as a precipitate under ethanol at -70°C for up to one month before use in reverse
 15 transcriptase (RT)-PCR.

ii. RT-PCR of *KIT* exon 16-19

First-strand cDNA synthesis was accomplished using the First-Strand cDNA Synthesis
 kit (Pharmacia Biotech) so that ~100 ng mRNA was randomly primed by 0.1 µg
 20 pd(N6) in a total volume of 15 µl. Two µl of the completed first cDNA strand reaction
 was then directly used per 12 µl PCR reaction by adding 10 µl PCR mix containing 10
 pmol each of the mouse/human derived primers KIT1F and KIT7R (5'-TCR TAC
 ATA GAA AGA GAY GTG ACT C and 5'-AGC CTT CCT TGA TCA TCT TGT
 AG, respectively; Moller et al. 1996, *supra*), 1.2 µl 10 x PCR-buffer (10 mM Tris-
 25 HCl, pH 8.3, 50 mM KCl) and 0.5 U of AmpliTaq polymerase (Perkin-Elmer)
 incubated with an equal amount Taqstart antibody (Clontech) at 25 °C for 5 min to
 achieve a hot start PCR. The reaction was covered with 20 µl mineral oil and
 thermocycled in a Hybaid Touchdown machine (Hybaid) with 40 cycles at 94°C for 1
 min, 55-48°C (touchdown one degree per cycle the first seven cycles and then 48°C in

the remaining cycles) for 1 min and 72°C for 1 min. After PCR 2µl loading dye was added to each sample which were then loaded on 4% agarose gel (Nusieve/Seakem 3:1, FMC Bioproducts) and electrophoresed with 100V for 80 min. Products were visualised by ethidium bromide staining and UV-illumination.

5

iii. Cloning and sequencing of RT-PCR-products

The RT-PCR products representing *KIT* exon 16-19 were purified by extraction from 2% agarose gels using the QIAEX gel extraction kit (QIAGEN) and cloned into the pUC18 vector using the Sureclone ligation kit (Pharmacia Biotech). Plasmids were isolated using the QIAFilter plasmid Midi kit (QIAGEN). Cloned plasmid inserts were sequenced using dye primer chemistry. Each cycling reaction was prepared with plasmid template DNA and ready reaction mix containing fluorescently labelled M13 forward or reverse primer as described in the ABI Prism protocol P/N 402113 (Perkin Elmer). Cycling and sample pooling were performed using a Catalyst 800 Molecular Biology Workstation (ABI) following the instruments user manual (Document number 903877, Perkin Elmer). The resulting extension products were purified, loaded and analysed using the 377 ABI Prism sequencer as described by the instrument protocol P/N 402078 (Perkin Elmer).

10

15

20

iv Results and discussion

25

A 424 bp fragment including *KIT* cDNA exon 16-19 was amplified from all pigs. The Hampshire pigs did not show any additional products whereas the Large White pigs (eight tested) all showed a 301 bp truncated cDNA fragment (Fig 2). Sequence analysis revealed the 424 bp fragment was identical in the two breeds whereas the whole exon 17 (123 bp) was missing from the 301 bp fragment. Apparent differences between individuals regarding the relative amounts of these two products may have been caused either by different genotypes containing differing numbers of copies of the *KIT* gene sequence, individual differences in mRNA expression levels or random RT-PCR effects.

The two upper fragments present in Large white pigs represent heteroduplexes between the 301 and 424 bp fragments (Fig. 2). This was shown by an experiment where these slow migrating fragments were generated by pooling homoduplexes of the 424 and 301 bp which were then heat denatured and cooled to 25°C. Moreover, cloning of the lower heteroduplex fraction of a Large White pig resulted in clones with insert length corresponding to either of the two homoduplexes.

Example 2

10 PCR Amplification and Sequencing of KIT Exon 17-Intron 17 (5' Splice Site)

i. PCR to produce DNA Sequencing Template

A 175 bp region including the boundary between exon17 and intron17 of the *KIT* gene was amplified for sequence analysis using forward primer KIT21 (5' - GTA TTC ACA GAG ACT TGG CGG C -3') and reverse primer KIT35 (5' - AAA CCT GCA AGG AAA ATC CTT CAC GG - 3'). PCR was carried out on a DNA thermal cycler (Perkin Elmer 9600) in a total volume of 20 µl containing 25 ng genomic DNA, 1.0 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 200 µM dNTPs, 0.5 U AmpliTaq Gold (Perkin Elmer) and 10 pmol of both KIT21 and KIT35 primer. To activate AmpliTaq Gold, initial heat denaturation was carried out at 94°C for 10 minutes followed by 32 cycles each consisting of 45 sec at 94°C, 45 sec at 55°C and 45 sec at 72°C. The final extension lasted for 7 min at 72°C. PCR products were cloned into vector pUC18 using the SureClone ligation kit (Pharmacia Biotech).

25 ii. Preparation of Plasmid DNA

Plasmid DNA was purified from overnight bacterial culture using the Jetstar plasmid midi kit (Genomed) and the resulting DNA diluted to 150 ng/µl.

iii. Sequencing of plasmid DNA

DNA was sequenced as in example 1 section iii.

iv. Results

5 A portion of the DNA sequence from exon 17 and intron 17 of the KIT gene was determined and compared between animals with each of these three alleles. Figure 3 shows that the *I* allele carries a splice site mutation at position 1 of intron 17. This G to A base substitution is present in one of the two gene copies carried on each chromosome. The base substitution occurs in the invariant GT dinucleotide which characterises 5' exon/intron boundaries. Analysis of the *I^P* allele showed the splice site mutation was not present in either the normal (KIT1) or duplicated copy of the gene (KIT2). We have found the splice site mutation is unique to the *I* alleles, and therefore makes it possible to distinguish the I-KIT2 sequences.

Example 3

15

Testing For the Presence of the Splice Site Mutation with PCR RFLP

20 To easily test for the presence of the G to A splice site mutation, restriction endonuclease *Nla*III (CATG) was used to exploit the point substitution identified at position 1 of intron 17 (Figure 3). The *Nla*III recognition sites in the fragment amplified from *KIT* and the expected restriction products are illustrated in Figure 4A and 4B respectively.

i. PCR to produce DNA for RFLP Test

25 The PCR to produce DNA for RFLP analysis was performed exactly as described in example 2.

ii. Restriction Enzyme Digestion and Electrophoresis

30 The PCR amplification product is 175 bp in length. To test for polymorphism at position 1 of intron 17, digestion reactions were set up as below:

3.0 µl PCR amplified DNA
 1.0 µl 10 X NEBuffer 4
 0.1 µl BSA 100 µg/ml
 0.1 µl *Nla*III 10 U/µl
 5.8 µl dH₂O

(1 X NEBuffer 4 (New England Biolabs) contains 50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate and 1 mM DTT). Following incubation at 37°C for 90 minutes each 10 µl reaction volume had 2 µl of loading dye added and the mix loaded on a 8% native polyacrylamide gel (Protogel, 37.5:1 acrylamide:bisacrylamide, National Diagnostics, Atlanta) in 0.5 X TBE (44.5 mM Tris pH 8.0, 44.5 mM boric acid and 0.5 mM EDTA) and electrophoresed for 3 hours at 200V in a vertical slab unit (SE600 Hoefer Scientific Instruments). Products were visualised by ethidium bromide staining.

iii Results

A PCR RFLP protocol was designed to test for the presence of the splice site mutation as the substitution occurs within the recognition site for restriction endonuclease *Nla*III. Figure 4B illustrates that presence of the G to A base substitution at position 1 of KIT intron 17 results in restriction at each of two *Nla*III recognition sites within the 175 bp DNA fragment. Following electrophoresis, this results in fragments of sizes 80 bp, 54 bp and 41 bp. Where the splice site mutation is absent however, incubation with *Nla*III results in digestion only at recognition site 1. Following electrophoresis this results in fragments of 134 bp and 41 bp. The invariant *Nla*III recognition site 1 serves as an internal control to ensure complete digestion has taken place. Results of this PCR RFLP analysis are illustrated in Figure 4C. Analysis was performed on fragments amplified from clones which either carry the splice site mutation (lane 2) or carry the normal splice site sequence (lane 3). Lane 4 shows the result of analysis where DNA amplified from the genomic DNA of a coloured animal was used. Lane 5 shows the resulting bands where a white animal was tested. The test was used to analyse 121

individuals from seven different breeds of pig. The splice site mutation was found only in the 97 animals with the dominant white phenotype ($I/-$ or I^*/i) and none of the 24 coloured (I^P or i) examples (Table 2). This analysis confirms I and I^* to be unique in that they are the only alleles to carry the splice site mutation.

5 **TABLE 2**
Distribution of the Splice Site Mutation Between Different Breeds and Coat Phenotype

Breed	Coat Colour	Assumed Genotype ¹	Animals Tested	Normally spliced KIT ²	Splice Mutation ²
Large White	white	$I/-$	33	33	33
Landrace	white	$I/-$	56	56	56
Hampshire	coloured	i/i	5	5	0
Duroc	coloured	i/i	5	5	0
Pietrain	coloured	i/i	8	8	0
Meishan	coloured	i/i	5	5	0
Wild Boar	coloured	i/i	1	1	0
Wild Boar x Large White	white	$I^*/-$	8	8	8
Totals					
	white	$I/-$	89	89	89
	white	$I^*/-$	8	8	8
	coloured	i/i	24	24	0

¹ White animals may be homozygous or heterozygous for the I allele

10 ² Presence of the splice site mutation determined by *Nla*III PCR RFLP test

Example 4

15 Quantification of Normal KIT and Splice Mutant KIT (Intron 17 nt1^{G→A})

As the splice site mutation is present in only one of the duplicated regions of I and not in the duplicated region of I^P , the various genotypes can be expected to have the attributes described in Table 3.

TABLE 3

Genotype	Copies of Normal KIT	Copies of KIT containing the splice mutation	Ratio of normal KIT to splice mutant KIT
I/I	2	2	1:1
I/i	2	1	2:1
i/i	2	0	2:0
I/I ^P	3	1	3:1
I ^P /i	3	0	3:0

Due to the dominance of allele *I*, three of the genotypes in Table 2 are carried by white animals and therefore can not be identified by phenotypic characterisation. Quantification of the relative amounts of the normal *KIT* gene and the splice mutant *KIT* gene allows the ratio between the two to be calculated, and therefore the genotype of individual animals predicted. This was achieved by quantification of two DNA fragments following *Nla*III digestion. The amount of 134 bp fragment, representative of the normally spliced *KIT* gene, and of 54 bp fragment, representative of the splice mutant *KIT*, were measured following electrophoresis using GeneScan software.

i. PCR to Produce DNA for Quantification

As described in example 2 section i. The reverse primer KIT35 is labelled with the ABI fluorescent dye FAM at the 5' end.

ii Restriction Enzyme Digestion

As described in example 2 section ii.

iii Electrophoresis and Quantification of DNA Fragments

5 Following digestion, 0.5 μ l of the reaction volume was mixed with 2.5 μ l of deionised formamide, 0.5 μ l of GS350 DNA standard (ABI) and 0.4 μ l blue dextran solution before being heated to 90°C for 2 minutes and rapidly cooled on ice. Three μ l of this mix was then loaded onto a 377 ABI Prism sequencer and the DNA fragments separated on a 6% polyacrylamide gel in 1 X TBE buffer for 2 hours at 700 V, 40 mA,
10 32 W. The peak area of fragments representative to both the normal and splice mutant forms of *KIT* were quantitated using the GeneScan (ABI) software.

iv. Ratio Calculations

The peak area value of the 134 bp fragment (normal *KIT*) was divided by twice the
15 peak area value of the 54 bp fragment (splice mutant *KIT*) in order to calculate the ratio value for each sample.

v. Results

Analysis was performed on animals from the Swedish wild pig/Large White intercross
20 pedigree for which genotypes at *I* have been determined by conventional breeding experiments with linked markers. Figure 5 and Table 4 show the ratio of normal to mutant *KIT* calculated for animals from each of the three genotype classes, *I/I* (expected ratio 1:1), *I/i* (expected ratio 2:1) and *I/I^P* (expected ratio 3:1). The results are entirely consistent with the expected ratio values and indicate that the three
25 genotype classes can be distinguished using this method.

TABLE 4

Ratio of the Two KIT Forms in Different Dominant White Genotypes in a Wild Pig/Large White Intercross

5

Genotype	Phenotype	Expected Ratio (Normal: Mutant)	Observed Ratio (Normal: Mutant) ± SE	Number Tested
I/I	white	1:1	1.15 ± 0.075	13
I/I ^P	white	3:1	3.11 ± 0.084	12
I/i	white	2:1	2.23 ± 0.109	14

Figure 5 illustrates that the range of ratio values calculated for the two genotypes *I/I* and *I/I^P* do not overlap. This enables animals carrying the *I^P* allele to be identified and the frequency of the allele within different pig breeds determined. Ratio values were calculated for 56 Landrace and 33 Large White animals and the results are shown in Figure 6. A clearly bimodal distribution is observed with 7 Landrace and 3 Large White individuals having a ratio value of approximately 3 or above, suggesting them to be heterozygous carriers for the *I^P* allele (genotype *I/I^P*). This means *I^P* has gene frequency estimates of 6.25% (7/112 chromosomes tested) and 4.5% (3/66 chromosomes tested) within the Landrace and Large White breeds respectively.

10

15

CLAIMS:

1. A method for determining coat colour genotype in a pig which comprises:
 - 5 (a) obtaining a sample of pig nucleic acid; and
 - (b) analysing the nucleic acid obtained in (a) to determine whether a mutation is present at one or more exon/intron splice sites of the *KIT* gene.
- ✓ 10 2. A method as claimed in claim 1 wherein the analysis in step (b) is carried out to determine whether a mutation is present at the exon 17/intron 17 boundary.
3. A method as claimed in claim 2 wherein the mutation consists of the substitution of the G of the conserved GT pair for A.
- 15 4. A method as claimed in any one of claims 1 to 3 wherein the sample of nucleic acid is amplified prior to analysis.
5. A method as claimed in claim 4 wherein the nucleic acid is genomic DNA.
- 20 6. A method as claimed in claim 5 wherein amplification is carried out using PCR and at least one pair of suitable primers.
7. A method as claimed in claim 6 wherein the pair of suitable primers is:
 - 25 5'-GTA TTC ACA GAG ACT TGG CGG C-3'); and
 - 5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3'.

8. A method as claimed in any one of claims 5 to 7 wherein after amplification the nucleic acid is treated with a restriction enzyme, followed by analysis of fragment lengths.

5 9. A method as claimed in claim 8 wherein the nucleic acid is treated with the restriction enzyme *Nla*III.

10. A method as claimed in claim 8 or claim 9 wherein the ratio of restriction fragment lengths is determined.

10

11. A method as claimed in claim 4 wherein the nucleic acid is mRNA.

12. A method as claimed in claim 11 wherein the nucleic acid is amplified using RT-PCR.

15

13. A method as claimed in claim 12 wherein the length of RT-PCR product is determined.

20

14. A method for determining coat colour genotype in a pig which comprises the step of analysing a sample of pig KIT protein to determine whether the protein is the splice variant protein.

25

15. A kit for use in determining the coat colour genotype of a pig which comprises one or more reagents suitable for determining whether a mutation is present at one or more exon/intron splice sites of the *KIT* gene.

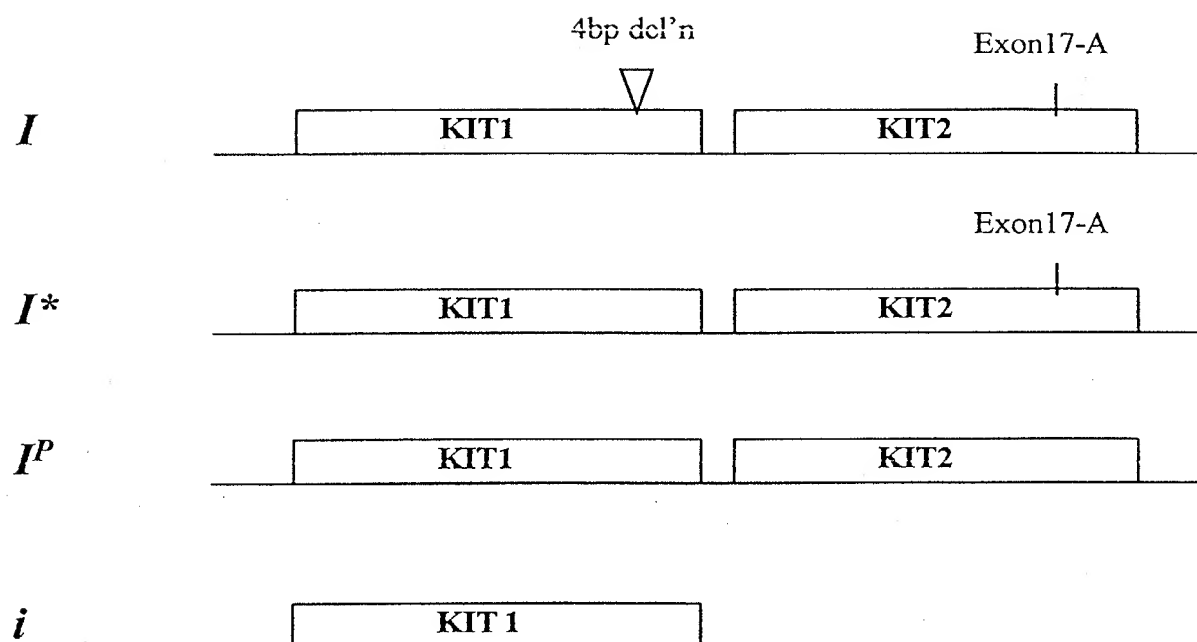
16. A kit as claimed in claim 15 which comprises one or more reagents for carrying out PCR and one or more pairs of suitable primers.

5 17. A kit as claimed in claim 16 which comprises the following pair of primers:

5'-GTA TTC ACA GAG ACT TGG CGG C-3'); and

5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3'.

Figure 1



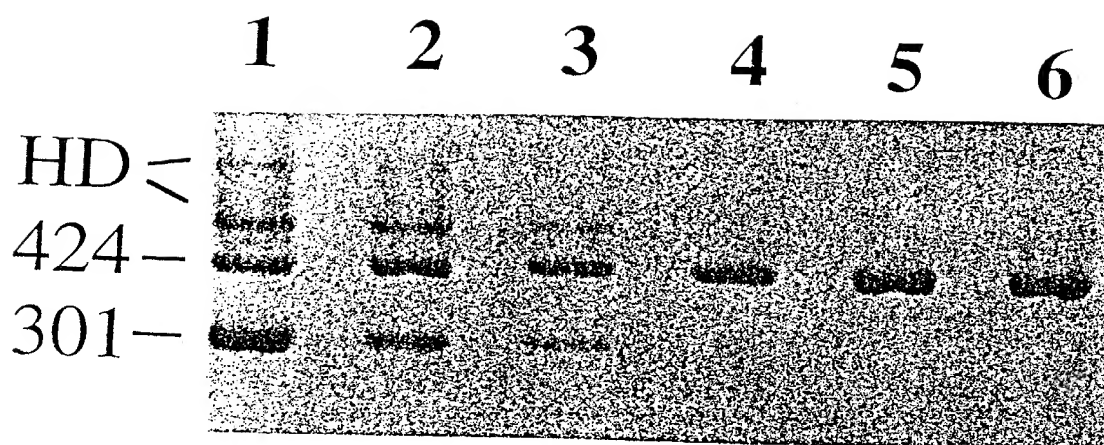


Figure 2

Figure 3

Sequence Alignment Across the Exon /Intron Border of KIT Exon 17

Allele	Gene Copy	Sequence	
		Exon 17	Intron 17
I	KIT1	AAT TAC GTG GTC AAA GGA AAC	↓ GTG AGT ACC CAC GCT CTC CTG ACA GTC
	KIT2 A... ..
I ^p	KIT1 G... ..
	KIT1 G... ..
i	KIT1 G... ..

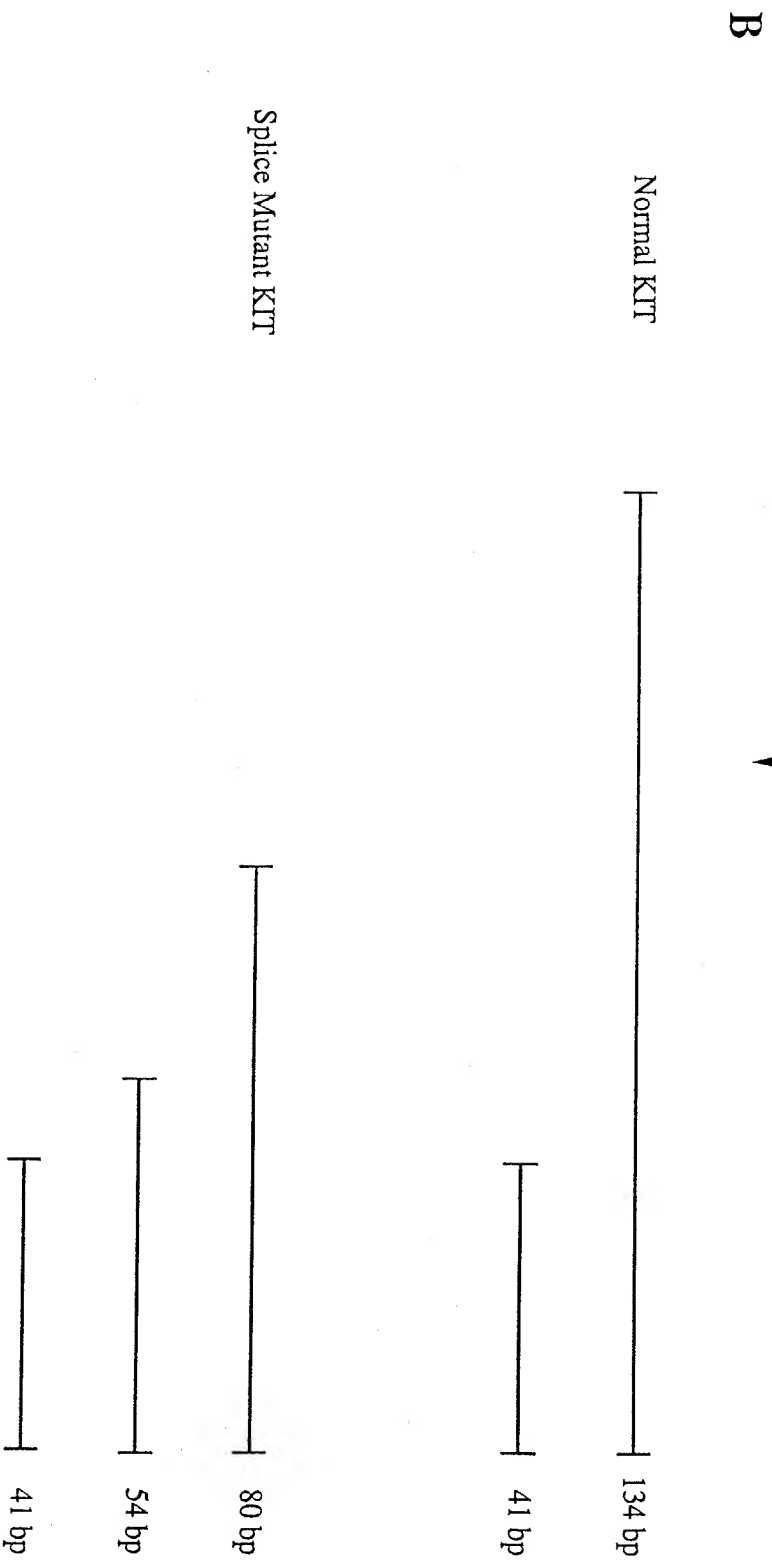
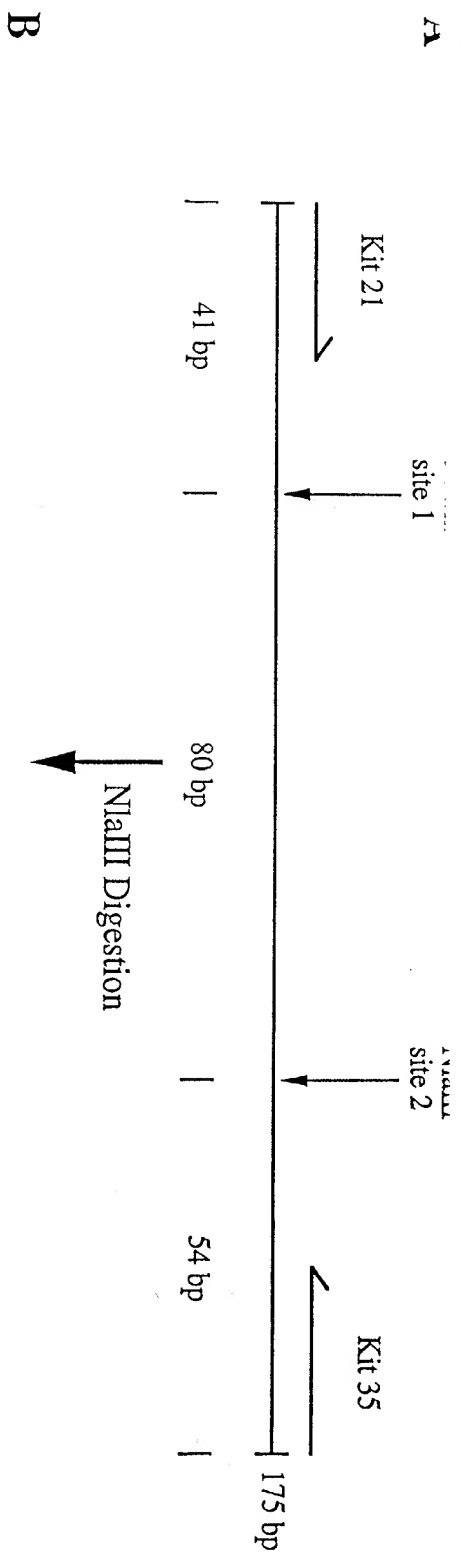


Figure 4 A+B

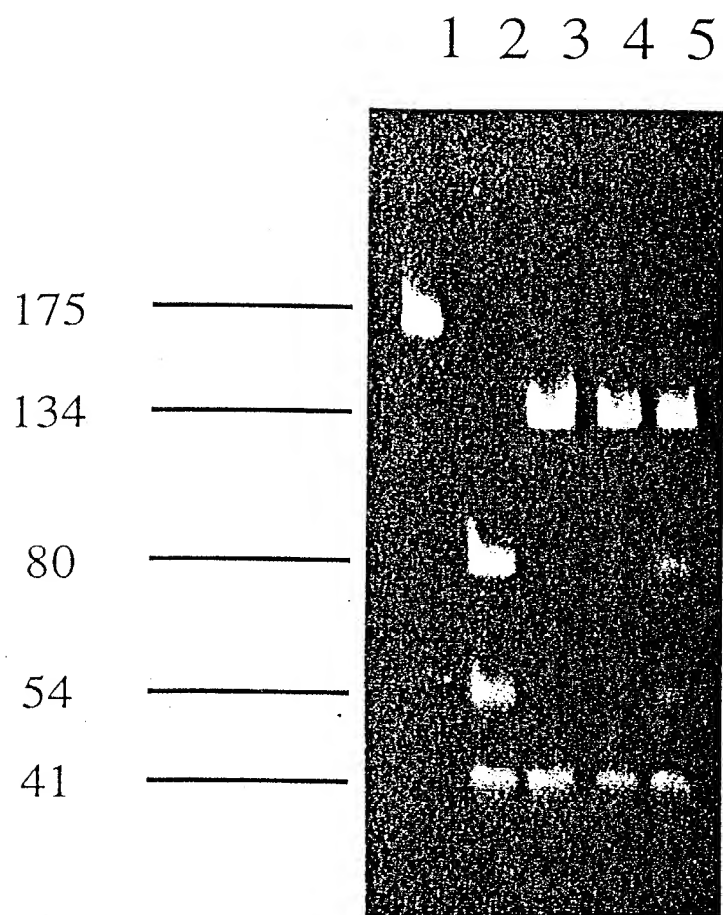


Figure 4C

Figure 5

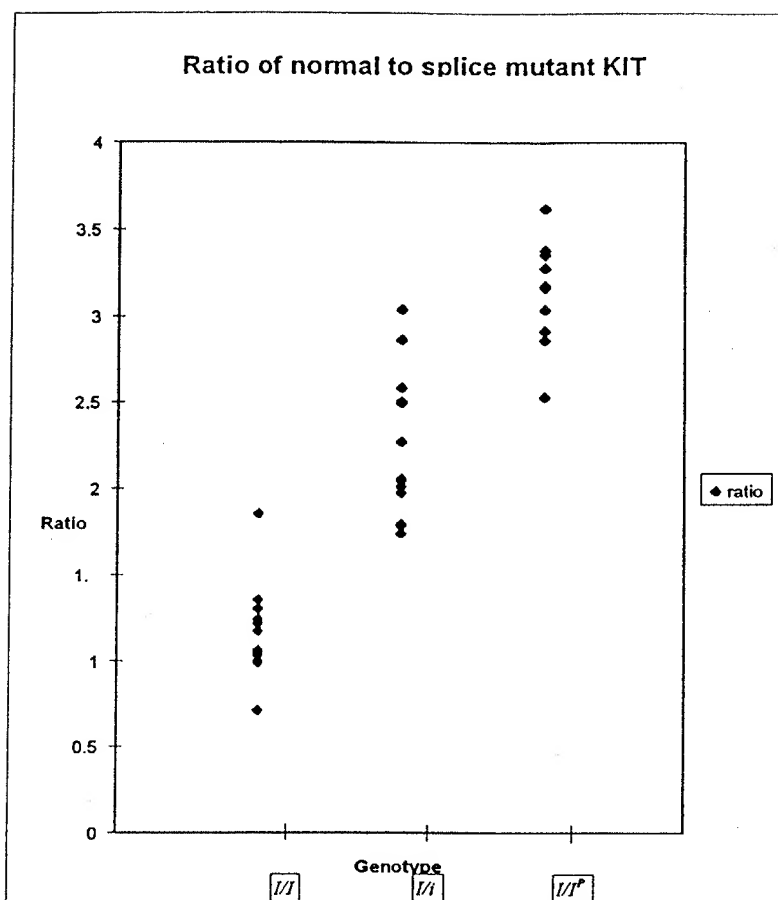


Figure 5 - Comparison of the ratio of normal to splice mutant KIT for three classes of genotype.

Figure 6

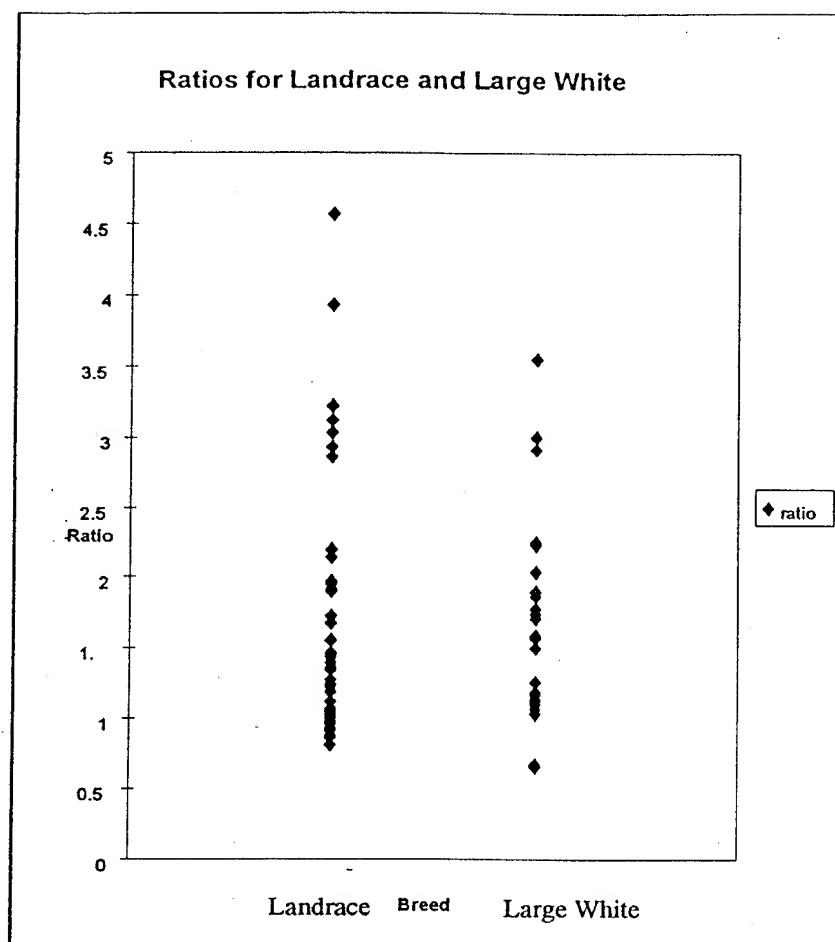


Figure 6 - Comparison of the ratio of normal to splice mutant KIT for two breeds of pig.